

ПРОГРАММНАЯ РЕАЛИЗАЦИЯ РАСЧЕТА НУКЛЕОТИДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ПРИ БИОСИНТЕЗЕ БЕЛКА

Трофименко Е.Г., Издебский В.А.

Общеизвестно, что генетический код – это определенные сочетания нуклеотидов и последовательность их расположения в молекуле ДНК. Это свойственный всем живым организмам способ кодирования аминокислотной последовательности белков при помощи последовательности нуклеотидов. Исследованием и моделированием структурно-функциональной организации различных генов занимаются при разработке современных лекарственных аппаратов, диагностических средств, проводников электричества и всевозможных наноструктур и наномеханических устройств, в криминалистике, в исследованиях молекулярной археологии и палеонтологии и др. Все это обуславливает актуальность создания удобных программных средств для анализа, исследования и моделирования функционального строения нуклеотидных последовательностей, кодирующих потенциальные белки.

Проанализировав существующие доступные в сети Интернет-сайты [1, 2], посвященные решению задачи расчета последовательностей нуклеотидов и аминокислот при биосинтезе белка было выявлено ряд характерных недостатков. К ним можно отнести, во-первых, отсутствие графического интерфейса для наглядности процесса; во-вторых, чрезмерное количество биологических терминов и отсутствие описаний и пояснений, что затрудняет понимание происходящих трансляций; в-третьих, отсутствие параллельного символического соответствия представления исходной последовательности аминокислот и результатов расчетов цепи. Кроме того, большинство существующих Интернет-ресурсов, посвященных проблемам генетики и генетическим исследованиям, занимаются торговлей биосырья и генетического материала и в их услуги не входит решение конкретной задачи расчета цепей биополимеров при биосинтезе белка. Помимо этого следует отметить, что предлагаемые программы, как правило, достаточно дороги для приобретения, а, следовательно, недоступны для широкого круга обычных пользователей.

Целью данной работы была разработка такой программного приложения, которое позволило бы производить расчет цепей нуклеиновых кислот и белковой молекулы на основных этапах биосинтеза белка при различных вариантах исходных данных, позволяло бы получить образное представление (отображение) пространственного соответствия кодонов и аминокислот в цепях, наглядное отображение возможных вариаций при изменении (или постоянстве) состава исходной цепи.

Молекула ДНК в наиболее типичной конфигурации представляет собой двойную закрученную плектонемически (т.е. требующую обязательного раскручивания при ее разделении на составные части) правую спираль, образованную двумя полинуклеотидными цепями (рис. 1). Пространственную структуру белка задаёт порядок расположения аминокислот. Количество аминокислот велико, но в белках встречается только 20 разных аминокислот. Их можно рассматри-

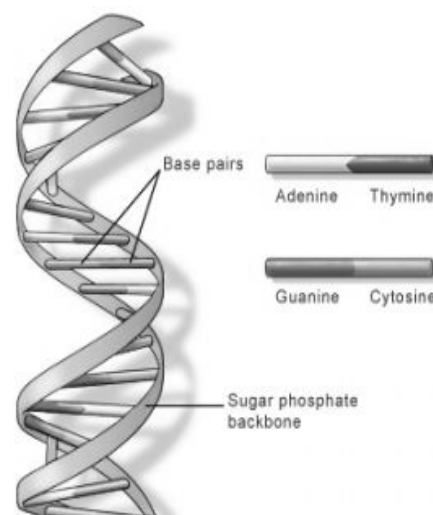


Рис. 1. Структура ДНК.

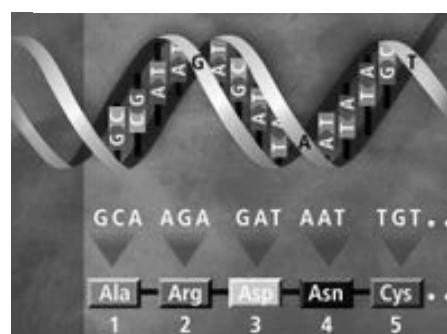


Рис. 2. Генетический код.

вать как алфавит «языка» белковой молекулы. Для удобства эти аминокислоты обозначают символами, используя одну или первые три буквы русского или английского названия аминокислоты, например аланин – Ала или А, глицин – Гли (Gly) или G [1]. Для кодирования 20 аминокислот, а также сигнала «стоп», означающего конец белковой последовательности, достаточно трёх последовательных нуклеотидов. Набор из трёх нуклеотидов называется триплетом. Принятые сокращения, соответствующие аминокислотам и кодонам, изображены на рис. 2.

В ДНК используется четыре нуклеотида – аденин (А), гуанин (G), цитозин (С), тимин (Т), которые обозначаются буквами А, G, С и Т [3]. Эти буквы составляют алфавит генетического кода. В РНК используются те же нуклеотиды, за исключением тимина, который заменён похожим нуклеотидом – урацилом, который обозначается буквой U. В молекулах ДНК и РНК нуклеотиды выстраиваются в цепочки и, таким образом, получают последовательности генетических букв. Причем, при формировании двойной спирали действует принцип комплементарности [4], согласно которому аденинам одной цепи соответствуют тимины другой, а гуанинам одной – цитозины другой и наоборот. Эти соотношения известны как правила Чаргаффа, который показал, что ДНК включает равные количества определенных азотистых оснований и вывел для двуцепочечной ДНК следующие закономерности: $A = T$, $G = C$; $A + G = C + T$; $A + C = G + T$. Все последующие работы с ДНК, независимо, молекулярная ли это биология либо другая, казалось бы, далекая от нее дисциплина, так или иначе построены на принципе комплементарности азотистых оснований.

Последовательность оснований в цепочке ДНК напрямую определяет последовательность оснований в РНК, на которую она «переписывается» в процессе, называемом транскрипцией. Транскрипция – процесс синтеза РНК с использованием ДНК в качестве матрицы, происходящий во всех живых клетках. Другими словами, это перенос генетической информации с ДНК на РНК (перекодировка).

Трансляция является финальной реакцией реализации генетической информации. Трансляция представляет собой осуществляемый рибосомой синтез белка из аминокислот на матрице информационной (или матричной) РНК (иРНК или мРНК).

На рис. 3 показан вид главного окна программы, меню которого позволяет выбрать и осуществить расчеты соответствующих цепей нуклеиновых кислот (рис. 4, а) и белковой молекулы (рис. 4, б).

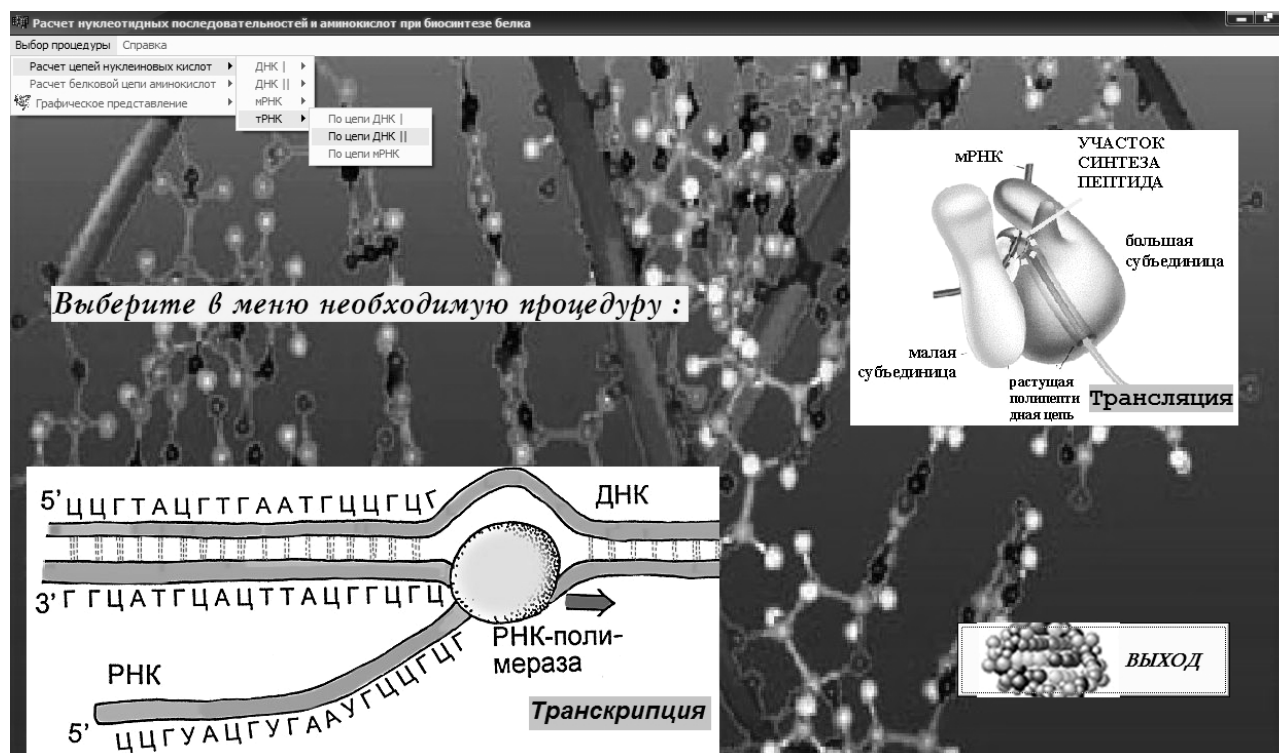
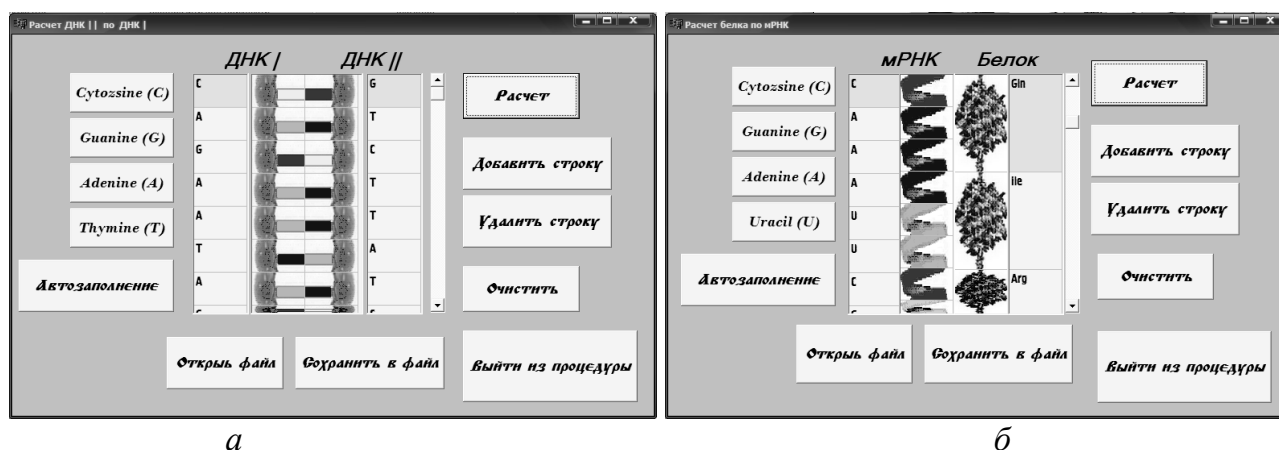


Рис. 3. Главное окно программы.



а

б

Рис. 4. Окна расчетов:

а) цепей нуклеиновых кислот; б) цепи аминокислот белка.

Разработанная программная реализация проста и непритязательна в использовании и позволяет вводить и редактировать нуклеотидные последовательности, транслировать нуклеотидную последовательность в белковую, транслировать белковую последовательность в нуклеотидную с учетом неравномерности использования кодонов-синонимов и представлять результаты в удобном (таблично-графичном) виде.

В разработанном программном приложении реализован простой и удобный пользовательский интерфейс средствами C++ Builder. Предусмотрена проверка корректности и соответствия введенных данных в диалоговом режиме. Программа предоставляет возможность сохранения вводимых и обрабатываемых данных в файл, а также выбор данных из ранее созданных файлов. Тем самым возможно создавать, анализировать и сохранять базы различных нуклеотидных последовательностей.

Также пользователю предоставлена возможность изучить и проанализировать

графическое представление основных этапов биосинтеза белка: транскрипции и трансляции (рис. 5).

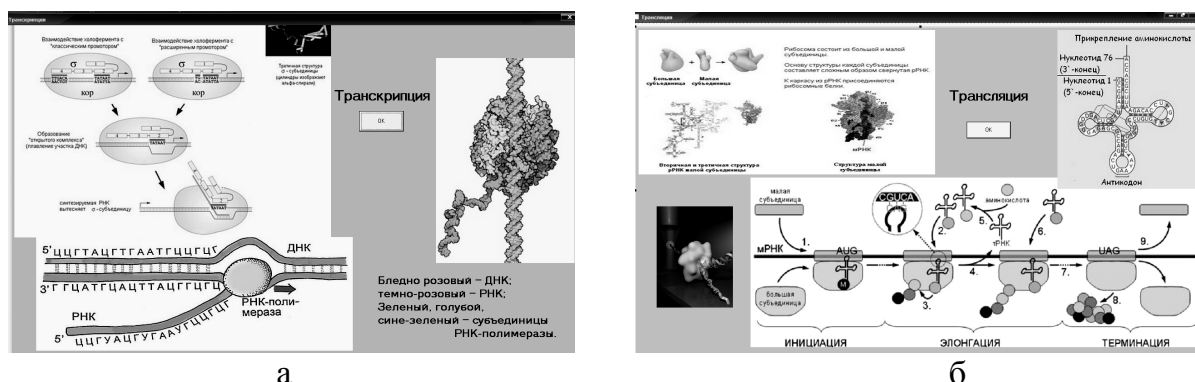


Рис. 5. Схематическое представление: а) транскрипции, б) трансляции.

Таким образом, предлагаемая программа отображает «математическую» сторону процесса биосинтеза белка и будет полезной для специалистов, различных сфер, использующих в своей работе принципы структурно-функциональной организации генов, а также научно-познавательной для обычного пользователя.

Литература

1. Режим доступа: <http://molbiol.ru/scripts/>.
2. Режим доступа: <http://oligos.ru/index.html>.
3. Чемерис А.В., Вахитов В.А. Новая старая ДНК. Уникальные черты самой главной молекулы или почему ученые разных специальностей в последнее время обращают на ДНК все больше внимания / Институт биохимии и генетики Уфимского научного центра РАН. – Уфа, 2002.
4. Шабарова З.А., Богданов А.А., Золотухин А.С. Химические основы генной инженерии. – М.: Изд-во МГУ, 1994.